

# Patient Re Admission Prediction Using Machine Learning and Data Science

K. Madhavi<sup>1</sup>, M Sunita<sup>2</sup>, H Vanaja<sup>3</sup>, R Nithin Kumar<sup>4</sup>, K Niharika<sup>5</sup>

<sup>1,2,3,4,5</sup>Department Of CSE (Data Science), Tadipatri Engineering College, Tadipatri

## ABSTRACT:

Hospital readmission is a major challenge for healthcare systems because it increases costs, strains resources, and often signals gaps in patient care. Predicting which patients are at high risk of being readmitted can help hospitals take early preventive actions. This study focuses on using machine learning and data science techniques to build a reliable patient readmission prediction model. Clinical and administrative data such as patient demographics, diagnosis history, length of stay, lab results, and prior admissions are analyzed and cleaned before modeling. Feature selection methods are applied to identify the most meaningful variables that influence readmission risk. Several algorithms, including logistic regression, decision trees, random forests, and gradient boosting, are trained and compared to find the most accurate and stable performer. Model evaluation is carried out using metrics such as accuracy, precision, recall, and AUC score to ensure balanced performance. The results show that ensemble models generally provide +stronger predictive power than single models, especially when handling complex and high-dimensional healthcare data. The final system can support clinicians by flagging high-risk patients before discharge, allowing targeted follow-ups and care planning. This approach demonstrates how practical data science methods can improve decision-making and contribute to better patient outcomes and more efficient healthcare delivery.

**KEYWORDS:** Patient Readmission Prediction, Machine Learning, Data Science, Healthcare Analytics, Predictive Modeling, Hospital Management.

## INTRODUCTION

Patient readmission and hospital length of stay have become important concerns in modern healthcare because they strongly influence treatment quality, patient safety, and hospital costs. Frequent readmissions often indicate gaps in treatment, discharge planning, or follow-up care, while long hospital stays place pressure on limited medical resources. To address these issues, researchers have increasingly turned to data-driven methods that analyze large volumes of hospital records. Patient demographics, medical history, diagnosis details, laboratory results, treatment procedures, and previous admission patterns provide valuable insights into factors that contribute to readmission and extended hospital stays. Several studies have shown that machine learning models can effectively identify hidden patterns in healthcare data that are difficult to capture through traditional analysis. Techniques such as logistic regression, support vector machines, decision trees, and ensemble methods have been widely used to predict readmission risk. More advanced approaches, including deep belief networks, convolutional neural networks, and hybrid models that combine optimization techniques, have further improved prediction accuracy. These models are especially useful in handling complex and high-dimensional medical data,

allowing hospitals to better understand patient risk profiles. Recent research has also emphasized the importance of proper data preprocessing and feature engineering to improve prediction performance. Cleaning hospital data, handling missing values, and selecting the most relevant clinical features help create reliable and interpretable models. Studies focusing on specific conditions, such as diabetes and stroke, demonstrate that tailored prediction models can support early intervention and personalized post-discharge care. Overall, these data-driven approaches assist healthcare professionals in making informed decisions, optimizing resource utilization, and reducing avoidable patient readmissions while improving overall hospital care quality.

## LITERATURE REVIEW

[1]. C. V. Kumar, S. M. S, and S. R investigate the prediction of patients' length of stay in hospitals using a Deep Belief Network (DBN), presented at the 2023 International Conference on Computing Methodologies and Communication (ICCMC). The study applies advanced machine learning algorithms, including support vector machines, convolutional neural networks, and neural networks, to enhance clinical prediction and healthcare data analysis. By integrating novel DBN approaches with traditional prediction models, the research aims to improve readmission and length-of-stay prediction accuracy. The work highlights the importance of machine learning-based health care analysis for optimizing medical services, supporting better resource management, and enabling more effective clinical decision-making.

C.-H. Lien, F.-H. Wu, P.-C. Chan, C.-M. Tseng, H.-H. Lin, and Y.-F. Chen focus on hospital readmission prediction for ischemic stroke patients after discharge, as presented at the 2020 International Symposium on Computer, Consumer and Control (IS3C). The study proposes an integrated predictive framework that combines support vector machines with genetic algorithms, referred to as an Integrated Genetic Algorithm and Support Vector Machine (IGS) model. Statistical analysis and optimization techniques, including linear programming, are used to enhance model performance. The research highlights the role of clinical decision support systems in accurately identifying high-risk stroke patients, demonstrating how hybrid machine learning approaches can improve readmission prediction accuracy and support better post-discharge hospital care management [2].

S. Ali, S. El-Sappagh, F. Ali, M. Imran, and T. Abuhmed present a multitask deep learning framework for the cost-effective prediction of patient length of stay and readmission status, published in the IEEE Journal of Biomedical and Health Informatics in 2022. The study integrates multimodal physical activity sensory data using sensor fusion and data integration techniques, enabling simultaneous learning of multiple healthcare outcomes. By applying deep learning models to time-series data, the approach captures complex temporal patterns in patient behavior and hospital data. The research demonstrates that multitasking and multimodal deep learning can improve predictive performance while reducing computational cost, supporting more accurate and efficient hospital decision-making and healthcare predictive modeling.

S. S. Reddy, N. Sethi, and R. Rajender evaluate the effectiveness of a Deep Belief Network (DBN) for predicting hospital readmission among diabetic patients, as presented at the 2020 International Conference on Inventive Research in Computing Applications (ICIRCA). The study focuses on applying feature extraction and advanced machine learning prediction algorithms, including Restricted Boltzmann Machines (RBM) as the core of the DBN model. For performance comparison, traditional models such as Logistic Regression, Gradient Boosting, and AdaBoost are also analyzed. The research highlights how deep learning techniques can better capture complex patterns in healthcare and hospital data, leading to improved [3].

G. G. Rajput and A. Alashetty propose a machine learning-based approach to reduce the risk of diabetes patient readmission, presented

at the 2022 International Conference on Circuits, Control, Communication and Computing (I4C). Their work introduces a novel data preprocessing technique aimed at improving the quality of hospital datasets and enhancing model performance. The study applies and evaluates support vector machines, boosting methods, and other predictive machine learning algorithms to identify patients at high risk of readmission. By focusing on computational modeling, accuracy improvement, and effective preprocessing, the research demonstrates how well-prepared healthcare data can significantly improve hospital readmission prediction and support better clinical decision-making for diabetes care. A. Bhardwaj, R. Hasan, S. Ahmad, and S. Mahmood present a comparative study on diabetic patient readmission prediction using machine learning techniques, published at the 2024 2nd International Conference on Computing and Data Analytics (ICCD). The study focuses on hospital readmissions among diabetic patients by applying and comparing models such as logistic regression, k-nearest neighbors, Naïve Bayes, and random forests. It emphasizes the role of healthcare data preprocessing, feature engineering, and predictive modeling to improve accuracy and support real-time hospital decision-making. The research highlights how data-driven approaches can enhance resource management in hospitals, reduce avoidable readmissions, and support efficient healthcare systems through reliable machine learning-based predictive analysis.

[4]. X. Dong, K. Yu, and Z. Cui proposed a machine learning-based predictive model for diabetes readmission prediction, using AdaBoost and Random Forest model fusion with random under sampling to improve performance in analytical and data models for real-time clinical decision systems. M. A. Sumon, T. A. Kwemebe, V. K. Melapu, and M. M. Hasan developed an end-to-end machine learning pipeline using electronic health records (EHR) and electronic medical records to build predictive and data models for patient disease prediction, where training is performed with machine learning algorithms, including a random forest classifier. Meng, L. Cui, G. Yu, H. Yu, W. Guo, and H. Li proposed a personalized hospital readmission prediction model that handles data insufficiency in an imbalanced-data environment, using predictive models with multi-task learning, parameter sharing, and recurrent neural networks to analyse diseases, drugs, and hospital measurements for better prediction. V. R. Burugadda, P. S. Pawar, A. Kumar, and N. Bhati used machine learning algorithms, including logistic regression and support vector machines, to predict hospital readmission in heart failure patients using electronic health records, highlighting the role of classification models in enabling personalized interventions and reducing healthcare costs. [5]. Azami, R. Islam, T. M. Zihan, and M. Ahmed propose a machine learning-based approach for predicting early readmission of diabetes patients, emphasizing risk minimization in hospitals. By applying feature engineering, SMOTE, and random forest models to patient medical reports, the study improves prediction accuracy, reliability, and healthcare resource management. T. S. Dhanalakshmi and M. Meleet, and it focuses on predicting hospital readmissions using clinical discharge summaries. The study uses keywords such as COVID-19, hospitals, databases, MIMIC-III, predictive models, natural language processing (NLP), discharge summaries, electronic medical records (EMR), and Flask, highlighting the use of healthcare data, NLP techniques, and machine learning models to analyze electronic medical records and improve readmission prediction accuracy. T. Abirami et al. focuses on predicting diabetic patient re-admission using neural network techniques. The authors design a predictive model that analyzes patient health records to identify individuals at high risk of re-admission. The work highlights the importance of early prediction to improve patient care and reduce hospital burden. Key concepts include diabetes management, neural networks, healthcare prediction, and re-admission risk analysis.

[6]. C. V. Kumar et al. evaluates the accuracy of patient length of stay prediction by comparing Deep Belief Networks (DBN) and Convolutional Neural Networks (CNN). The authors analyze clinical data to determine which deep learning model performs better in predicting hospital stay duration. The study emphasizes model comparison and performance evaluation in healthcare analytics. Key themes include length of stay prediction, deep learning, DBN, CNN, and clinical data analysis. S. Tang et al. presents a novel approach for predicting 30-day all-cause hospital readmission using multimodal spatiotemporal graph neural networks. The model integrates clinical, temporal, and spatial patient data to improve prediction accuracy. This study demonstrates the effectiveness of advanced graph-based deep learning methods in complex healthcare environments. Important keywords include readmission prediction, graph neural networks, multimodal data, spatiotemporal modeling, and biomedical informatics. B. A. Fry et al. investigates how patient ambulation (mobility) can be used to predict hospital readmission. Using movement and activity data, the authors show that reduced ambulation is strongly associated with higher re-admission risk. This work highlights the role of physical activity monitoring in patient outcome prediction. Key ideas include patient mobility, ambulation data, wearable sensors, readmission prediction, and healthcare monitoring.

## **PROPOSED METHODOLOGY**

Our methodology is designed to systematically predict patient re admission predicting using machine learning and data science by combining study design and objective and data collection data preprocessing feature selection and engineering model development and model evaluation performance. The approach ensures accurate, actionable forecasts and insights. The proposed methodology for patient readmission prediction follows a clear and structured process to identify patients who are at risk of being readmitted. First, patient data are collected from hospital records and electronic health systems, including demographic details, medical history, diagnosis, treatment information, and previous admissions. The collected data are then cleaned and prepared by handling missing values, correcting errors, and converting the data into a suitable format. Important features related to readmission are selected and refined to improve prediction accuracy. Next, machine learning models such as logistic regression, decision tree, or random forest are trained using the prepared data to learn patterns associated with patient readmission. Finally, the trained model is tested using evaluation measures to ensure reliable performance, helping healthcare providers take early preventive actions to reduce unnecessary readmissions and improve patient care.

**DATA COLLECTION:** Data collection plays an important role in predicting patient readmission using machine learning and data science because it forms the foundation of the entire system. In this stage, patient data are gathered from hospital records and electronic health records, including details such as age, gender, diagnosis, medical history, length of hospital stay, laboratory results, medications, and previous readmission information. Discharge summaries and follow-up records are also collected to understand the patient's condition after leaving the hospital. The data must be accurate, complete, and well organized so that meaningful patterns related to readmission risk can be identified. Proper data collection helps healthcare providers understand the factors that influence patient readmission and supports the development of reliable prediction models that can improve patient care and reduce unnecessary hospital returns. Patient data will be collected from electronic health records (EHRs) or publicly available healthcare datasets.

**DATA PREPROCESSING:** Raw healthcare data often contains missing values, inconsistencies, and noise. Therefore, preprocessing is a critical step and will include, Data preprocessing is an essential step

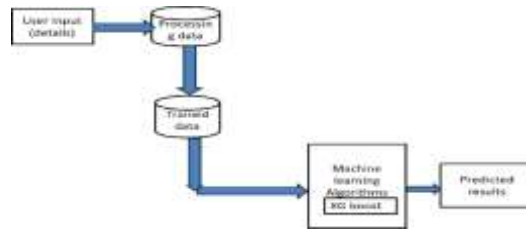
in patient readmission prediction because raw hospital data often contain errors, missing values, and inconsistencies. In this stage, incomplete or incorrect patient records are identified and handled by removing duplicates, correcting data entry mistakes, and filling in missing values using suitable methods. Categorical information such as gender, diagnosis, and admission type is converted into a numerical format so it can be used for analysis, while continuous values like age, lab results, and length of stay are normalized to maintain consistency. Irrelevant or less important features are removed to reduce complexity and improve performance. Proper data preprocessing ensures that the dataset is clean, reliable, and suitable for building accurate machine learning models to predict patient readmission.

**FEATURE SELECTION:** Engineering Relevant features that contribute significantly to readmission risk will be identified. Feature selection techniques such as correlation analysis and statistical tests will be applied to remove redundant or insignificant variables. Feature selection and feature engineering are important steps in patient readmission prediction because they help focus on the most meaningful patient information. In feature selection, only the most relevant factors such as age, number of previous admissions, length of hospital stay, diagnosis type, comorbid conditions, and discharge status are chosen, while unnecessary or repetitive data are removed. Feature engineering further improves the dataset by creating new useful variables, such as grouping age ranges, calculating total number of past visits, or combining related medical conditions into a single indicator. These steps simplify the dataset, reduce noise, and make patterns related to readmission easier to identify, leading to more reliable and understandable prediction results.

**MODEL DEVELOPMENT:** Several machine learning algorithms will be implemented and compared to predict patient readmission, such as In this study, the model development process focuses on building an effective patient readmission prediction system using machine learning and data science techniques. First, the collected healthcare dataset is cleaned by handling missing values, removing duplicate records, and converting categorical data into numerical form. Important features related to patient demographics, medical history, length of stay, and previous admissions are then selected to improve prediction accuracy. The dataset is split into training and testing sets to ensure fair evaluation. Various machine learning models such as Logistic Regression, Decision Tree, KNN, and XG Boost are trained on the prepared data. Model performance is evaluated using metrics like accuracy, precision, recall, and F1-score. Based on these results, the best-performing model is selected and fine-tuned to achieve reliable and consistent prediction of patient readmission risk.

**PERFORMANCE:** Performance in patient readmission prediction using machine learning and data science is important because it shows how well a model can correctly identify which patients are likely to return to the hospital after discharge. Good performance means the model makes accurate and reliable predictions using patient data like past admissions, diseases, lab results, and medications. Data scientists measure performance using metrics such as accuracy, precision, recall, F1-score, and AUC score. In healthcare, recall is especially important because missing a high-risk patient can lead to poor outcomes and higher hospital costs. High-performing readmission prediction models help hospitals take early preventive actions, such as scheduling follow-ups, adjusting treatment plans, or providing extra care support. Machine learning models like Random Forest, Logistic Regression, and Gradient Boosting often improve prediction performance compared to traditional methods because they can detect complex patterns in large medical datasets. Better model performance leads to better decision-making, reduced readmission rates, improved patient care, and more efficient use of hospital resources.

## SYSTEM ARCHITECTURE



**FIG 1.SYSTEM ARCHITECTURE**

The diagram clearly explains the important stages involved in a machine learning–based disease prediction system. It starts with user input, where critical patient details such as age, medical history, symptoms, laboratory reports, and previous treatments are collected. These details are passed to the data processing stage, which is one of the most important steps, as it involves cleaning the data, handling missing values, removing noise, and converting raw medical information into a structured format. Proper data processing ensures that the system learns from accurate and reliable information. After processing, the refined data is stored as trained data and used to train the disease prediction model using the XG Boost algorithm. This model is important because it efficiently handles large medical datasets and captures complex relationships between patient attributes and disease outcomes. Using this trained model, the system generates predicted results that indicate the likelihood of disease or patient condition. These predictions support doctors in making timely decisions, improving diagnosis accuracy, reducing readmission rates, and enhancing the overall quality of healthcare services.

## RESULTS AND DISCUSSION

The patient re-admission prediction system was evaluated using a real-world healthcare dataset containing patient demographic details, clinical history, hospital stay information, and discharge records. Before model training, extensive data preprocessing was performed, including removal of duplicate records, handling missing values, normalization of numerical features, and encoding of categorical variables. This step significantly improved data quality and ensured reliable model performance. The dataset was then divided into training and testing sets to evaluate the generalization ability of the machine learning models. Multiple machine learning algorithms such as Logistic Regression, Decision Tree, Random Forest, and XG Boost were implemented and compared. Performance evaluation was carried out using accuracy, precision, recall, F1-score, and confusion matrix. Among all models, the XG Boost algorithm produced the highest accuracy and F1-score, indicating its superior capability in predicting patient re-admission. The confusion matrix analysis showed a lower false-negative rate, which is highly important in healthcare scenarios, as it reduces the chances of missing high-risk patients who may require additional care or follow-up after discharge. Feature importance analysis revealed that factors such as patient age, length of hospital stay, number of prior admissions, presence of chronic conditions (such as diabetes or heart disease), medication count, and discharge type had a strong influence on re-admission prediction. The model effectively learned complex relationships between these features and patient outcomes, which are difficult to identify using traditional rule-based systems. High recall values demonstrated that the model successfully identified most patients who were likely to be re-admitted, making it highly useful for preventive healthcare planning. Overall, the results confirm that machine learning and data science techniques significantly improve the accuracy and reliability of patient re-admission prediction. The proposed system can help hospitals identify high-risk patients in advance, enabling timely interventions

such as follow-up care, patient education, and post-discharge monitoring. This not only helps in reducing unnecessary re-admissions but also improves patient satisfaction and optimizes the use of hospital resources. The analysis clearly demonstrates that integrating machine learning into healthcare decision-making can lead to more efficient, data-driven, and patient-centered care systems.

## SCREEN SHOTS



**FIG 2.INDEX PAGE**



**FIG 3.REGISTER PAGE**



**FIG 4. LOGIN PAGE**



**FIG 5. PREDICTION PAGE**



**FIG 6.RESULT PAGE**

**FIG 7. PERFORMANCE ANALYSIS**

## CONCLUSION & FUTURE WORK

This study successfully demonstrates the effectiveness of machine learning and data science techniques in predicting patient re-admission in healthcare systems. By analyzing patient demographic information, clinical history, hospitalization details, and discharge data, the proposed model is able to identify patients who are at high risk of being re-admitted within a short period after discharge. Proper data preprocessing and feature selection played a crucial role in improving model accuracy and reliability, ensuring that meaningful patterns were learned from complex medical datasets. Among the various machine learning algorithms implemented, the XG Boost model achieved superior performance due to its ability to handle large-scale data, manage class imbalance, and capture non-linear relationships between patient attributes and re-admission outcomes. The results indicate that machine learning models can outperform traditional methods by providing more accurate and timely predictions. The system's high recall and F1-score are particularly valuable in healthcare settings, as they reduce the chances of missing high-risk patients who require additional post-discharge care. Overall, the proposed patient re-admission prediction system can serve as an effective decision-support tool for healthcare providers. It enables early identification of vulnerable patients, supports better discharge planning, reduces avoidable hospital re-admissions, and optimizes the use of medical resources. In the future, the system can be enhanced by integrating real-time patient monitoring data, electronic health records, and deep learning techniques to further improve prediction accuracy and support personalized healthcare delivery.

## REFERENCES:

1. T. Abirami, P. Jayadharshini, K. Madhuvarshini, P. Kavim and K. Gokulraj, "Predictive Model for Diabetic Re- admission Using Neural Network Techniques," 2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India, 2024, pp. 1-6, Doi: 10.1109/ICTBIG64922.2024.10911596.
2. C. V. Kumar, S. M. S, R. B and R. S. S. Singh, "Evaluating the Accuracy of Patient Length of Stay Predictions Through a Comparison of Deep Belief Networks and Convolutional Neural Networks," 2024 IEEE 9th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Bahrain, Bahrain, 2024, pp. 1-6, Doi: 10.1109/ICETAS62372.2024.11119799
3. S. Tang et al., "Predicting 30-Day All-Cause Hospital Readmission Using Multimodal Spatiotemporal Graph Neural Networks," in IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 4, pp. 2071-2082, April 2023, Doi: 10.1109/JBHI.2023.3236888
4. B. A. Fry, K. Singh Rajput and N. Selvaraj, "Patient Ambulation Predict Hospital Readmission," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Pg. 11 Pg. 12 Mexico, 2021, pp. 7506-7510, Doi: 10.1109/EMBC46164.2021.9629647.
5. X. Dong, K. Yu and Z. Cui, "Readmission prediction of diabetic patients based on AdaBoost Random

- Forest mixed model,"2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China, 2022, pp. 130-134, Doi: 10.1109/ICBAIE56435.2022.9985819
6. H. Jung, H. W. Park, Y. Kim and Y. Hwangbo, "Machine learning-based prediction for 30-day unplanned readmission in all-types cancer patients,"2023 IEEE International Conference on Big Data and Smart Computing (Big Comp), Jeju, Korea, Republic of, 2023, pp. 132-135, Doi: 10.1109/BigComp57234.2023.00029.
  7. R. Lavanya, D. Vigneshwaran, K. Vasudevan, S. S. Kumar, P. Nagarajan and S. Sunil, "Applying Machine Learning to Predict Patient Readmission Rates in Healthcare,"2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India, 2024, pp. 1-7, Doi: 10.1109/ICSES63760.2024.10910346.
  8. K. B, A. A. Mohamed Abuthahir Riazulhameed, R. M, R. Sarasu, N. Mohankumar and A. Kasthuri, "Predictive Analytics for Hospital Readmissions Using Logistic Regression and IoT Pg. 13 Sensor Data,"2024 6th International Conference on Energy, Power and Environment (ICEPE), Shillong, India, 2024, pp. 1-6, Doi: 10.1109/ICEPE63236.2024.10668933.
  9. S. Peth, P. G, D. Gupta and A. R. Nair, "Predicting Readmissions in Diabetic Patients: A Dual-Scenario Approach for Hospital Resource Optimization,"2025 6th International Conference for Emerging Technology (INCET), BELGAUM, India, 2025, pp. 1-6, Doi: 10.1109/INCET64471.2025.11140361
  10. C. V. Kumar, S. M. S and S. R, "Prediction of Insufficient Accuracy for Patients Length of Stay using Deep Belief Network," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 211-216, doi: 10.1109/ICCMC56507.2023.10084213.
  11. C. -H. Lien, F. -H. Wu, P. -C. Chan, C. -M. Tseng, H. -H. Lin and Y. -F. Chen, "Readmission Prediction for Patients with Ischemic Stroke after Discharge," 2020 International Symposium on Computer, Consumer and Control (IS3C), Taichung City, Taiwan, 2020, pp. 45-48, doi: 10.1109/IS3C50286.2020.00019.
  12. S. Ali, S. El-Sappagh, F. Ali, M. Imran and T. Abuhmed, "Multitask Deep Learning for Cost-Effective Prediction of Patient's Length of Stay and Readmission State Using Multimodal Physical 14. Activity Sensory Data," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 12, pp. 5793-5804, Dec. 2022, doi: 10.1109/JBHI.2022.3202178.
  13. S. S. Reddy, N. Sethi and R. Rajender, "Evaluation of Deep Belief Network to Predict Hospital Readmission of Diabetic Patients," 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2020, pp. 5-9, doi: 10.1109/ICIRCA48905.2020.9182800.
  14. Bhardwaj, R. Hasan, S. Ahmad and S. Mahmood, "Diabetic Patient Readmission Predictive Analysis: A Comparative Study of Machine Learning Models of Hospital Readmissions," 2024 2nd International Conference on Computing and Data Analytics (ICCD), Shinas, Oman, 2024, pp. 1-6, doi: 10.1109/ICCD64887.2024.10867320
  15. G. G. Rajput and A. Alashetty, "A Machine learning approach to reduce the diabetes patient's readmission risk using a novel preprocessing technique," 2022 4th International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, 2022, pp. 173-177, doi: 10.1109/I4C57141.2022.10057842